General Clique Percolation in Network Evolution

Jingfang Fan[1] and Xiaosong Chen[1,*]

[1] *State Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, P.O. Box 2735, Beijing 100190, China*

We introduce a general (k, l) clique community, which consists of adjacent k-cliques sharing at least l vertices with $k - 1 \geq l \geq 1$. The emergence of a giant (k, l) clique community indicates a (k, l) clique percolation, which is studied by the largest size gap Δ of the largest clique community during network evolution and the corresponding evolution step T_c. For a clique percolation, the averages of Δ and T_c and the root-mean-squares of their fluctuations have power law finite-size effects whose exponents are related to the critical exponents. The fluctuation distribution functions of Δ and T_c follow a finite-size scaling form. In the evolution of the Erdős-Rényi network, there are a series of (k, l) clique percolation with $(k, l) = (2, 1), (3, 1), (3, 2), (4, 1), (4, 2), (5, 1), (4, 3)$, and so on. The critical exponents of clique percolation depend on l, but are independent of k. The universality class of a (k, l) clique percolation is characterized alone by l.

PACS numbers: 02.10.Ox, 89.75.Hc, 05.70.Fh, 64.60.-i
Keywords:

Community structure is of great interest in the studies of networks [1, 2]. The term network community is defined as a group of vertices that are more densely connected each other than other vertices in a network. The clique [3] and core [4] are the two examples of network community. To analyze the overlapping community structure of networks, Palla *et al.*[5] proposed the clique percolation method (CMP) to build up the communities from k-cliques, which is a fully connected subset of k vertices. Two k-cliques are considered to be adjacent if they share $k - 1$ vertices. A clique community is defined as the maximal union of k-cliques that can be reached from each through a series of adjacent k-cliques.

When the size of a clique community is comparable to network size N, a giant clique community emerges and there is a clique percolation. Derényi *et al.* [6] studied the clique percolation of the Erdős-Rényi (ER) model[7]. The k-cliques percolation takes place when the probability of connecting two vertices in the network reaches the threshold $p_c(k) = [(k - 1)N]^{-1/(k-1)}$ [6]. The normal percolation transition of ER model corresponds to the $k = 2$ clique percolation.

In this Letter, we introduce a general (k, l) clique community, where any two adjacent k-cliques share at least l vertices with $k - 1 \geq l \geq 1$. For the same network, the clique communities of different (k, l) are different. As an illustration, $(3, 1)$ and $(3, 2)$ clique communities in a network are shown in Fig.1 and they are different. A (k, l) clique percolation appears with the emergence of a giant (k, l) clique community. The previous clique percolation discussed in Ref.[6] corresponds to the special case $l = k - 1$ of general (k, l) clique percolation. The normal percolation transition corresponds to a $(2, 1)$ clique percolation. Here we study the general (k, l) clique percolation of the ER model by analyzing the finite-size effects of network evolution. From the power law exponents of finite-size effects, we can obtain the critical exponents of (k, l) clique percolation.

At first, we calculate the threshold of the (k, l) clique percolation using the approach developed by Newman, Watts, and Strogatz[9]. From the probability distribution p_k of vertex degrees, we define the generating functions $G_0(x) = \sum_{k=0}^{\infty} p_k x^k$ and $G_1(x) = \sum_{k=0}^{\infty} q_k x^k$, where $q_k = p_{k+1}(k + 1)/ < k >$ is the excess degree distribution of network [10].

The generating function $H_1(x)$ of the probability distribution for the sizes of the components reached from a randomly chosen edge satisfies a self-consistent equation

$$H_1(x) = xG_1(H_1(x)). \tag{1}$$

When its first derivative $H_1'(1)$ becomes infinite, a giant component appears and there is a percolation. This corresponds to $z' = G_1'(1) = 1$.

After taking k-clique as the new unit and considering two cliques to be adjacent if they share at least l vertices, the average degree of a k-cliques is

$$z = \sum_{k'=l}^{k-1} \binom{N-k}{k-k'} \binom{k}{k'} p^{\binom{k}{2}-\binom{k'}{2}}, \tag{2}$$

where p is the probability to connect two vertices with an edge. Correspondingly, the average excess degree of the k-

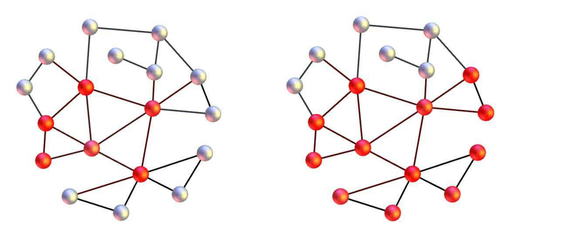

FIG. 1: (color online) Sketches of different clique communities in a network. The vertices belonging to a clique community are marked with red color. The largest $(3, 2)$ clique community is shown on the left and the largest $(3, 1)$ clique community on the right.

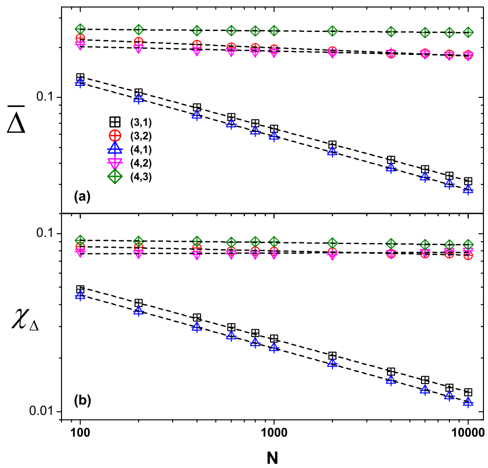

FIG. 2: (color online) Log-log plot of the average size gap $\bar{\Delta}$ and the root-mean-square of its fluctuation χ_Δ versus N for different (k,l) clique percolation. Power law behaviors of $\bar{\Delta}$ and χ_Δ are confirmed by simulation data. From the slopes of fitting lines, β_1 and β_2 can be obtained and are summarized in Table I.

cliques is

$$z' = \sum_{k'=l}^{k-1} \binom{N-k}{k-k'} \left[\binom{k}{k'} - 1\right] p^{\binom{k}{2}-\binom{k'}{2}}. \quad (3)$$

The threshold $p_c(k,l)$ of (k,l) clique percolation can be determined from the equation

$$\sum_{k'=l}^{k-1} \frac{\binom{k}{k'}-1}{(k-k')!} \left[1+O(N^{-1})\right] \left[N p_c^{\frac{k+k'-1}{2}}\right]^{k-k'} = 1\ , \quad (4)$$

which gives

$$p_c(k,l) = a(k,l) N^{-\frac{2}{k+l-1}} \left[1 + O(N^{-\frac{k-l-1}{k+l-1}})\right], \quad (5)$$

$$a(k,l) = \left[\frac{\binom{k}{l}-1}{(k-l)!}\right]^{-\frac{2}{(k-l)(k+l-1)}}. \quad (6)$$

For $l = k-1$, the correction term in Eq.5 vanishes and $p_c(k,k-1) = [(k-1)N]^{-\frac{1}{k-1}}$, which is in agreement with the result of Derényi et al [6, 11]. For $l < k-1$, the correction term $O(N^{-\frac{k-l-1}{k+l-1}})$ exists. For general (k,l), only $p_c(k,l) \sim N^{-\frac{2}{k+l-1}}$ is obtained by Bollobás and Riordan[8].

In the following, we study (k,l) clique percolation of ER model with Monte Carlo simulation and finite-size effects of network evolution [14]. During an evolution process, the largest (k,l) clique community has a size gap $S_1(T) - S_1(T-1)$ at an evolution step T. The largest reduced size gap during the whole evolution process is

$$\Delta \equiv \frac{1}{N} max\left(\{S_1(T) - S_1(T-1)\}\right), \quad (7)$$

which could be related to a percolation transition at the corresponding evolution step T_c. From the results of Δ and T_c in many Monte Carlo simulations, we can calculate the average size gap $\bar{\Delta}$ and the average transition point $\bar{T}_c$. Since $T = p \times N(N-1)/2$ and therefore $\bar{T}_c \propto N^{2-\frac{2}{k+l-1}}$, it is convenient to introduce a reduced evolution step $r = T/N^{2-\frac{2}{k+l-1}}$ and the reduced transition point $r_c \equiv T_c/N^{2-\frac{2}{k+l-1}}$. We anticipate that $\bar{\Delta}$ and $\bar{r}_c$ have the power law finite-size effects as

$$\bar{\Delta}(N) \sim N^{-\beta_1}, \quad (8)$$

$$\bar{r}_c(N) - r_c(\infty) \sim N^{-1/\nu_1}. \quad (9)$$

Using $p_c(k,l)$ in Eq. 5, we get

$$r_c^T(\infty) = \frac{1}{2} a(k,l). \quad (10)$$

The character of clique percolation is determined by the exponent β_1. The clique percolation is continuous when $0 < \beta_1 < 1$ and discontinuous when $\beta_1 = 0$ [15].

The fluctuations $\delta\Delta = \Delta - \bar{\Delta}(N)$ and $\delta r_c = r_c - \bar{r}_c(N)$ are investigated also. Their root-mean-squares are defined as

$$\chi_\Delta = \sqrt{\langle[\delta\Delta]^2\rangle}, \quad (11)$$

$$\chi_r = \sqrt{\langle[\delta r_c]^2\rangle}, \quad (12)$$

which decay algebraically as

$$\chi_\Delta \sim N^{-\beta_2}, \quad (13)$$

$$\chi_r \sim N^{-1/\nu_2}. \quad (14)$$

We anticipate a finite-size scaling form of fluctuation distribution functions as

$$P_\Delta(\delta\Delta, N) = N^{\beta_2} f_1(\delta\Delta N^{\beta_2}), \quad (15)$$

$$P_r(\delta r_c, N) = N^{1/\nu_2} f_2(\delta r_c N^{1/\nu_2}). \quad (16)$$

The universality class of continuous clique percolation is characterized by the critical exponents β_1, β_2, ν_1 and ν_2. Different clique percolation with the same critical exponents belong to the same universality class.

In Fig.2(a), the average $\bar{\Delta}$ is plotted with respect to network size N for $(k,l) = (3,1), (3,2), (4,1), (4,2)$, and $(4,3)$. The log-log plot of $\bar{\Delta}$ versus N show that $\bar{\Delta} \propto N^{-\beta_1}$. From the slope of fitting line, we can get the exponent β_1. It has been obtained that $\beta_1 = 0.331(5)$ for $(2,1)$, $\beta_1 = 0.32(2)$ for $(3,1)$, $\beta_1 = 0.33(1)$ for $(4,1)$, and $\beta_1 = 0.33(3)$ for $(5,1)$. Within error bars, β_1 for $l = 1$ and different k are equal. At $l = 2$, we get $\beta_1 = 0.05(2)$ for $(3,2)$ and $\beta_1 = 0.04(2)$ for $(4,2)$. The exponents β_1 of $l = 2$ are different from that of $l = 1$. At $l = 3$, $\beta_1 = 0.01(1)$ for $(4,3)$ and is different from that of $l = 1$ and $l = 2$. So the universality class of (k,l) clique percolation is characterized by l. In Fig.2(b), the log-log plot of χ_Δ versus N is shown for different (k,l). The slope of curve gives the exponent β_2. It is found that β_2 is equal to the corresponding β_1 within error bars.

The finite-size effects of $\bar{r}_c(N) - r_c(\infty)$ are shown in Fig. 3(a) for $l < k-1$ and Fig. 3(b) for $l = k-1$. From

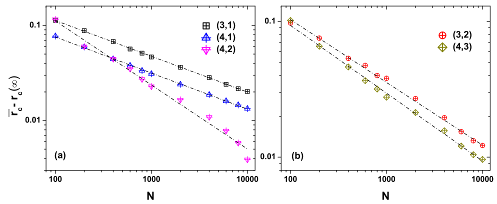

FIG. 3: (color online) Log-log plot of $\bar{r}_c - r_c(\infty)$ versus N. The results of $l < k-1$ is shown in (a) and $l = k-1$ in (b). The slopes of fitting lines give $1/\nu_1$, which is defined in Eq. 9 and summarized in Table I.

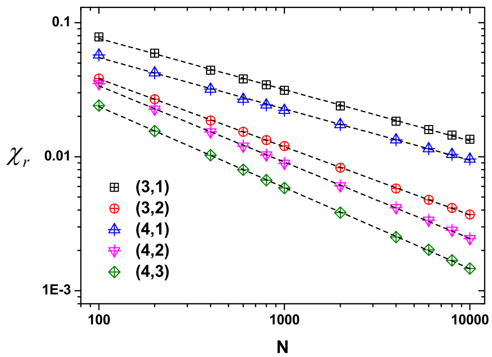

FIG. 4: (color online) Log-log plot of χ_r versus N. The slopes of fitting lines give $1/\nu_2$, which is defined in Eq. 12 and summarized in Table I.

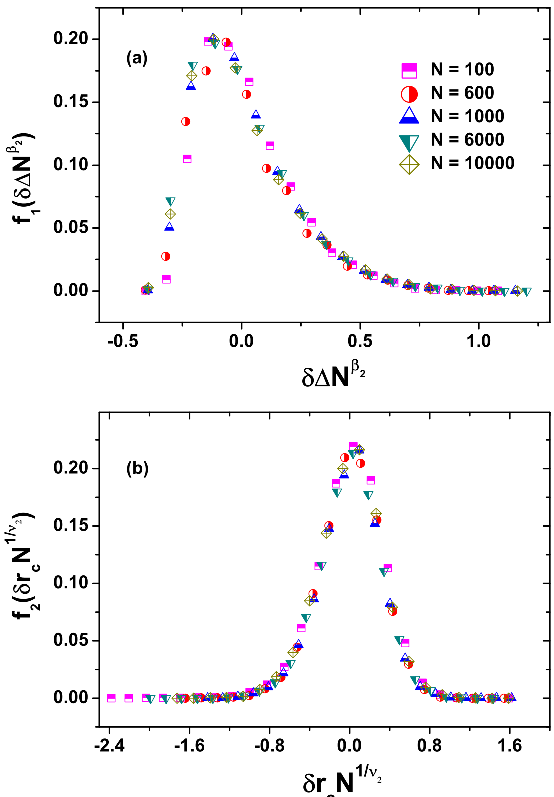

FIG. 5: (color online) Finite-size scaling functions of fluctuation distribution functions (a) $P_\Delta(\delta\Delta, N)$ and (b) $P_r(\delta r_c, N)$ for $(4,1)$ clique percolation.

the Monte Carlo data, the values of $r_c(\infty)$ and ν_1 are determined simultaneously. For $l = k-1$, we get $1/\nu_1 = 0.46(2)$ for $(3,2)$ and $1/\nu_1 = 0.49(4)$ for $(4,3)$. For $l < k-1$, we obtain $1/\nu_1 = 0.37(2)$ for $(3,1)$, $1/\nu_1 = 0.38(2)$ for $(4,1)$, and $1/\nu_1 = 0.67(20)$ for $(4,2)$. The $\bar{r}_c(N)$ of $(4,2)$ clique percolation shows deviation from a simple power law of N. This deviation could be resulted by additional finite-size terms. $1/\nu_1$ of different k and the same l are equal within error bars. We summarize $r_c(\infty)$ and $1/\nu_1$ of different (k,l) clique percolation in Table I.

The root-mean-square of δr_c is shown in Fig.4. Our results of simulations confirm the power-law behavior of χ_r. For $k = 3$, we get $1/\nu_2 = 0.33(5)$ for $l = 1$ and $1/\nu_2 = 0.50(2)$ for $l = 2$. At $k = 4$, we obtain $1/\nu_2 = 0.35(3)$ for $l = 1$, $1/\nu_2 = 0.52(4)$ for $l = 2$, and $1/\nu_2 = 0.60(3)$ for $l = 3$. Other results of $1/\nu_2$ are given in Table I. $1/\nu_2$ of different k and the same l are equal within error bars. This confirms further that the universality of (k,l) clique percolation is characterized alone by l. The exponent ν_2 agrees with ν_1 at $l = 1, 2$, but differs from ν_1 at $l = 3$.

For a network with size N, we simulate its evolution for many times. At each simulation, we can get the largest reduced size gap Δ and the corresponding transition point r_c. From the results of many simulations, the fluctuation distribution functions $P_\Delta(\delta\Delta, N)$ and $P_r(\delta r_c, N)$ can be obtained. Five different network sizes $N = 100, 600, 1000, 6000, 10000$ are chosen. Each fluctuation distribution function has five different curves. Using the finite-size scaling forms of Eq. 15 and 16, five curves of each distribution function collapse into one curve of finite-size scaling function. The scaling variables are defined with the exponents β_2 and ν_2 obtained from χ_Δ and χ_r. In Fig. 5, the finite-size scaling functions $f_1(\delta\Delta N^{\beta_2})$ and $f_2(\delta r_c N^{1/\nu_2})$ are shown for $(4,1)$ clique percolation. f_1 and f_2 of other (k,l) clique percolation have similar behavior.

In the evolution process of the ER model beginning from N isolated nodes to being fully connected with $N(N-1)/2$ vertices, there are a series of (k,l) clique percolation. The transition point of a (k,l) clique percolation is $T_c = \frac{1}{2}a(k,l)N^{2-\frac{2}{k+l-1}}$, which is of the same order of N for different clique percolation with equal $k+1$. The sequence of clique percolation in the ER model is as following: a) $(2,1)$ clique percolation at $T_c = 0.5N$, b) $(3,1)$ clique percolation at $T_c = 0.5N^{4/3}$, c) $(3,2)$ clique percolation at $T_c = 2^{-3/2}N^{3/2} \simeq 0.354N^{3/2}$, d) $(4,1)$ clique percolation at $T_c = 2^{-5/6}N^{3/2} \simeq 0.561N^{3/2}$, e) $(4,2)$ clique percolation at $T_c = 2^{-1}2.5^{-1/5}N^{8/5} \simeq 0.416N^{8/5}$, f) $(5,1)$ clique percolation at $T_c = 2^{-1}6^{1/10}N^{8/5} \simeq 0.598N^{8/5}$, g) $(4,3)$ clique percolation at $T_c = 2^{-1}3^{-1/3}N^{5/3} \simeq 0.347N^{5/3}$, and so on.

For comparison with the investigations above, we investigate directly the size of the largest (k,l) clique community in a network to study general clique percolation. For a network

TABLE I: Summary of reduced transition points $r_c = T_c/N^{2-\frac{2}{k+l-1}}$ and critical exponents of (k,l) clique percolation. $r_c(\infty)$ is obtained from Eq. 9 and the analytic result $r_c^T(\infty)$ is given in Eq. 10. The results of $(2,1)$ clique percolation are taken from[14].

(k,l)	$r_c(\infty)$	$r_c^T(\infty)$	β_1	β_2	$1/\nu_1$	$1/\nu_2$
(2,1)	0.5006(7)	2^{-1}	0.331(5)	0.334(2)	0.331(7)	0.334(3)
(3,1)	0.50(1)	2^{-1}	0.32(2)	0.36(4)	0.37(2)	0.33(5)
(3,2)	0.34(2)	$2^{-3/2}$	0.05(2)	0.03(2)	0.46(2)	0.50(2)
(4,1)	0.56(1)	$2^{-5/6}$	0.33(1)	0.35(2)	0.38(2)	0.35(3)
(4,2)	0.40(1)	$2^{-1}2.5^{-1/5}$	0.04(2)	0.03(2)	0.67(20)	0.52(4)
(4,3)	0.34(2)	$2^{-1}3^{-1/3}$	0.01(1)	0.01(1)	0.49(4)	0.60(3)
(5,1)	0.60(1)	$2^{-1}6^{1/10}$	0.33(3)	0.38(4)	0.30(5)	0.37(6)

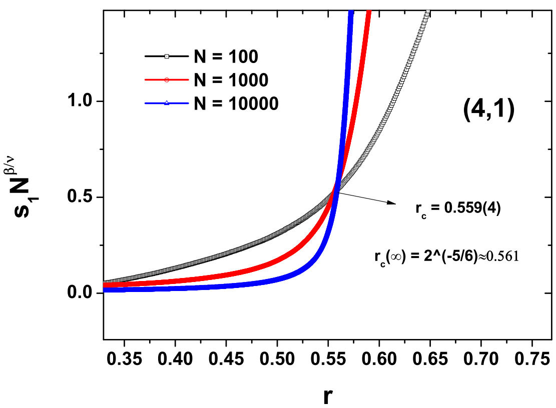

FIG. 6: (color online) Scaling size of the largest $(4,1)$ clique community $s_1N^{\beta/\nu}$ versus the reduced edge number $r = T/N^{3/2}$. We take $\beta/\nu = 0.33$. There is a fixed point at $r_c = 0.559(4)$.

with N vertices and T edges, the reduced size of the largest (k,l) clique community is denoted as $s_1(T,N)$. Near the transition point T_c of a (k,l) clique percolation, we anticipate a finite-size scaling form

$$s_1(T,N) = N^{-\beta/\nu}\widetilde{s}_1(tN^{1/\nu}), \tag{17}$$

where $t = (T-T_c)/T_c$ and ν is the critical exponent of correlation length. In the limit $N \to \infty$, $s_1(T,\infty) = 0$ at $T < T_c$ and $s_1(T,\infty) = A_1 t^\beta$ at $T > T_c$. Our previous investigations [12, 13] has confirmed the finite-size scaling form Eq. 17 for $(2,1)$ clique percolation.

At the transition point T_c, $s_1(T,N)N^{\beta/\nu}|_{T=T_c} = \widetilde{s}_1(0)$ and is independent of network size N. Using this property, we can determine the transition point T_c of (k,l) clique percolation from the fixed point of $s_1(T,N)N^{\beta/\nu}$.

In Fig. 6, we plot $s_1(T,N)N^{\beta/\nu}$ of $(4,1)$ clique community as a function of r for different N. $\beta/\nu = 0.33$ has been taken. We find a fixed point at $r_c = 0.559(4)$, which agrees with $r_c(\infty) = 0.56(1)$ obtained from network evolution. We expect that β/ν is equal to β_2 [14].

In summary, we introduce a (k,l) clique community consisting of adjacent k-cliques sharing at least l vertices with $k-1 \geq l \geq 1$. There is a (k,l) clique percolation with the emergence of a giant (k,l) clique community. We study the (k,l) clique percolation by investigating the largest reduced size gap Δ of the largest clique community during network evolution and the corresponding evolution step T_c. If the average size gap $\bar{\Delta} \sim N^{-\beta_1}$, there is a continuous clique percolation for $0 < \beta_1 < 1$ and a discontinuous clique percolation when $\beta_1 = 0$. The reduced transition point $\bar{r}_c$ obtained from T_c has the finite-size effect $\bar{r}_c - r_c(\infty) \sim N^{-1/\nu_1}$. The values of $r_c(\infty)$ obtained from Monte Carlo data agree with the analytic result of Eq. 5, which is derived using generating function method. The sequence of (k,l) clique percolation in the ER model is as following: $(2,1), (3,1), (3,2), (4,1), (4,2), (5,1), (4,3)$, and so on. The root-mean-squares of fluctuations $\delta\Delta$ and δr_c are found to decay algebraically as $\chi_\Delta \sim N^{-\beta_2}$ and $\chi_r \sim N^{-1/\nu_2}$. The fluctuation distribution functions follow the finite-size scaling forms $P_\Delta(\delta\Delta,N) = N^{\beta_2}f_1(\delta\Delta N^{\beta_2})$ and $P_r(\delta r_c,N) = N^{1/\nu_2}f_2(\delta r_c N^{1/\nu_2})$, respectively.

The universality of (k,l) clique percolation is characterized by the critical exponents β_1, β_2, ν_1 and ν_2. It has been found that the critical exponents of (k,l) clique percolation are independent of k, but dependent on l. The universality class of (k,l) clique percolation is characterized alone by l.

This work is supported by the National Natural Science Foundation of China under grant 11121403.

* Electronic address: chenxs@itp.ac.cn

[1] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).
[2] S. Fortunato, Phys. Rep. **486**, 75 (2010).
[3] B. Bollobás, *Random graphs*, 2nd ed. (Cambridge University Press, Cambridge, 2001).
[4] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes , Phys. Rev. Lett. **96**, 040601 (2006).
[5] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature (London) **435**, 814 (2005).
[6] I. Derényi, G. Palla, and T. Vicsek, Phys. Rev. Lett. **94**, 160202 (2005).
[7] P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).
[8] B. Bollobás and O. Riordan, Random Structures & Algorithms. **35**, 3 (2009)
[9] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[10] M. E. J. Newman, *Networks: an introduction*, (Oxford University Press, 2010), p. 445.
[11] G, Palla, I. Derényi, and T. Vicsek, J. Stat. Phys, **128**, 1 (2007).
[12] J.F. Fan, M.X. Liu, L.S. Li, and X.S. Chen, Phys. Rev. E **85**, 061110 (2012).
[13] M.X. Liu, J.F. Fan, L.S. Li, and X.S. Chen, Eur. Phys. J. B **85**, 132 (2012).
[14] Jingfang Fan, Maoxin Liu, Yong Zhu, and Xiaosong Chen, Finite-size Scaling in Phase Transition of Network Evolution, to be submitted.
[15] J. Nagler, A. Levina, and M. Timme, Nature Phys. **7**, 265 (2011).